ABSTRACT
This paper describes the systems developed by the LIUM laboratory for the 2009 WMT evaluation. We participated in the Arabic and Chinese/English BTEC tasks. We developed three different systems: a statistical phrase-based system using the Moses toolkit, an Statistical Post-Editing (SPE) system and a hierarchical phrase-based system based on Joshua. A continuous space language model was deployed to improve the modeling of the target languages. These systems are combined by a confusion network based approach.

INTRODUCTION
New features with respect to last year’s system:
• Arabic/English BTEC task and first participation in the Chinese/English BTEC track.
• Bitexts and LM resources limited to provided BTEC data
• Classical SMT system based on Moses
• First experiments with Joshua-based hierarchical system
• Development of a statistical post-editing system (SPE)
⇒ First steps in system combination

RESOURCES
Characteristics of the provided BTEC data

<table>
<thead>
<tr>
<th>corpus</th>
<th>Fr</th>
<th>En</th>
<th>Ar</th>
<th>Zh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>499</td>
<td>3744</td>
<td>3069</td>
<td>1</td>
</tr>
<tr>
<td>Dev1</td>
<td>506</td>
<td>3703</td>
<td>17.7k</td>
<td>16</td>
</tr>
<tr>
<td>Dev2</td>
<td>500</td>
<td>3900</td>
<td>17.8k</td>
<td>16</td>
</tr>
<tr>
<td>Dev3</td>
<td>506</td>
<td>3801</td>
<td>19.2k</td>
<td>16</td>
</tr>
<tr>
<td>Dev4</td>
<td>489</td>
<td>3612</td>
<td>16.5k</td>
<td>6</td>
</tr>
<tr>
<td>Dev5</td>
<td>500</td>
<td>3931</td>
<td>17.4k</td>
<td>16</td>
</tr>
</tbody>
</table>

• Training on Bnc + Dev1-3
• Development on Dev6, internal test on Dev7
• For some systems, Dev6 was added to the training material after tuning, keeping all settings unmodified.
• The Arabic texts were tokenized using SYSTRAN’s sentence analysis module. It includes a morphological decomposition.
• Chinese characters were segmented using tools from SYSTRAN
• All models are case-sensitive and with punctuations

SYSTEM ARCHITECTURE
SMT system
• Statistical phrase-based system using Moses and own tools
• Two pass approach
  • Generation of 1000-best lists with 4-gram back-off LM
  • Rescoring of those n-best lists with continuous space LM

SPE System
• Use of an SMT system to correct the errors of a rule-based system
  [Dugas et al, WMT’07, Simard et al, WMT’07]
• Here: SYSTRAN version 6 + Moses
• The LM is identical to the one used in the stand-alone SMT system

SYSTEM COMBINATION
The system combination approach is based on confusion network decoding similar to previous work:
1. 1-best hypotheses from all 3 systems are aligned using TER and confusion networks are built.
2. All confusion networks are connected into a single lattice. (the initial version of our system did not use weights for each branch)
3. A 4-gram LM is used to decode the resulting lattice and the best hypothesis is generated.

Scores and parameters
• Loïc: we need some details here (1st and final version)

EXPERIMENTAL EVALUATION
Case-sensitive BLEU scores of all the systems

<table>
<thead>
<tr>
<th>Approach</th>
<th>Arabic/English</th>
<th>Chinese/English</th>
<th>SPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMT + SPE CSLM</td>
<td>50.35</td>
<td>36.04</td>
<td>-</td>
</tr>
<tr>
<td>SMT CSLM</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SPE CSLM</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SMT + Hierarchical + manual tuning</td>
<td>+1.14</td>
<td>+0.84</td>
<td>-</td>
</tr>
<tr>
<td>Narrow tuning</td>
<td>+0.25</td>
<td>+0.65</td>
<td>-</td>
</tr>
<tr>
<td>Improvement of 1.6 BLEU on Ar/En (instead of 0.5 BLEU)</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Loïc: have you done it for Zh/En ??</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSION AND PERSPECTIVES
• Development of three complementary systems: Moses, Joshua and SPE
• But the SPE showed a very good generalization behavior on the eval data.
• Continuous space LM was useful as in previous experiments
• Initial work in system combination

Ongoing work
• Continuous space LM
• Unsupervised training on news-trainfr
• System combination

ACKNOWLEDGMENTS
This work has been partially funded by the French Government under the project DATA (ANR [C5006_143038]) and the European Commission under the project EuromatrixPlus (FP7 ICT 231720).